Hunter Fihn

CSC 4950

**WikiScrape Documentation**

<https://github.com/hunterfihn/WikiScrape-Submission>

The GUI was built using CustomTkinter, an extension and add-on to Python’s native Tkinter library. Support and documentation for CustomTkinter can be found here:

<https://customtkinter.tomschimansky.com/documentation>

Tkinter itself was used in order to handle the file dialog window for exporting the selected sources, and the Pillow package was used for image processing. More information can be found here:

<https://docs.python.org/3/library/dialog.html>

Image handling, in order to be consistent across machines, is pathed using OS and system operations in the **functions.py** file. When the executable is packaged, including the images in the packaging process allows the user to have the appropriate favicon and “no link found” image without needing to download anything separate. \*

OS, SYS:

<https://docs.python.org/3/library/os.html#module-os>

<https://docs.python.org/3/library/sys.html#module-sys>

The process of visiting the target Wikipedia page and web-scraping the sources section(s) is done with two libraries. For requesting the URL and retrieving the target page, the Python Requests library was utilized. For the HTML web-scraping, BeautifulSoup4 (BS4) was used. URL generation for Requests was completed by hardcoding the beginning of all Wikipedia pages, and then concatenating the user’s search term at the end.

Requests, BS4

<https://pypi.org/project/requests/>

<https://pypi.org/project/beautifulsoup4/>

The Google Gemini API was used for AI summaries of selected sources. The source code will not include my personal API key, but users can obtain one for free the link below. Gemini (at the time of writing this) is free to use when not exceeding the token threshold. When the key has been obtained, the user can copy/paste their key into the **functions.py** file between the quotes of the *googAPIkey* variable found below the imported libraries. \*\*

<https://ai.google.dev/gemini-api/docs/api-key>

All libraries can be installed via the pip installer. However, to make setup easier, there is a text document that can be used to mass install the required libraries. There are some unused libraries in this document, but I have yet to see any performance issues as a result. The document is the **requirements.txt** document, and the user can mass install the libraries with the following terminal command:

pip install -r requirements.txt

That should cover everything to get one started. Feel free to play around/modify the code in any way you choose. Messing around with the AI summary response may be fun, although maybe not as practical for the actual intended function. Enjoy WikiScrape!

\* Note: There are only two images, the favicon and the no link found image, both of which can be found in the repository’s files in the “Images” folder

\*\* Note: No WikiScrape.exe file is included in the WikiScrape GitHub Repository, as users need to supply their own Google Gemini API key before packaging the executable. When the key has been added into the source code, the following line can be run in the terminal to package the executable:

pyinstaller --onefile --add-data "Images/\*;images" --icon=Images/Favicon.ico --noconsole WikiScrape.py